

# **The Use of Cognitive Interviewing to Study Cultural Variation in Survey Response**

**Gordon Willis and Deirdre Lawrence**

*National Cancer Institute, National Institutes of Health*

**Martha Kudela and Kerry Levin**

*Westat*

**Kristen Miller**

*National Center for Health Statistics, Centers for Disease Control and Prevention*

Methods for the development, translation, and pretesting of survey questions across cultures, and across language, are undergoing significant evolution (Behling & Law, 2000; de la Puente & Pan, 2003; Gerber, 1999; Harkness, Van de Vijver, & Mohler, 2003; Ji, Schwarz, & Nisbett, 2000; Johnson, 1998; McKay, Breslow, Sangster, Gabbard, Reynolds, Nakamoto, & Tarnai, 1996; Rose & de la Puente, 2003; Stewart & Napoles-Singer, 2000; Warnecke, Johnson, Chavez, Sudman, O'Rourke, Lacey, & Horm, 1997). In particular, researchers are increasingly interested in determining the usefulness of pretesting and evaluation methods such as cognitive interviewing, behavior coding (see Cannell, Fowler, & Marquis, 1968; Fowler & Cannell, 1996), and respondent debriefing (Willis, 2005). It is not a simple matter, however, to apply these methods, as extension to multiple languages poses particular challenges in terms of staffing, analysis, and interpretation of results. The current paper will summarize two projects that involve the application of cognitive interviewing. The first of these was concerned with the details of *conducting* cross-cultural interviews; the second focused on the manner in which *results can be systematically processed* in a way that facilitates cross-cultural evaluation.

## **1. Study 1. The Current Population Survey Tobacco Use Supplement: Evaluation of Asian Language Translations**

A full description is provided by Kudela, Levin, Tseng, Hum, Lee, Wong, McNutt, & Lawrence (2004). In brief, the project required the evaluation of an instrument on tobacco use, originally developed in English, as translated to (a) Chinese (Mandarin and Cantonese), (b) Korean, and (c) Vietnamese. The following sections describe each stage of the project, with mention of difficulties that arose, and solutions that were enacted.

### **1.1. Staff selection**

Because Westat staff responsible for pretesting spoke no Asian languages, they created the position of "Survey Language Consultant" (SLC) for each of the three languages (the Chinese SLC was fluent in both Mandarin and Cantonese), to oversee cognitive interviewing activities, to contribute cultural insights relevant to the testing effort, to provide translation services as needed throughout the project, and to translate recruiting, data collection, and training materials. Qualified SLCs were not numerous, and personal contacts produced the best candidates (e.g. the wife of a Westat employee's husband's friend who was a social worker who heads a local Vietnamese agency). Although SLCs varied, as far as being local or at a distant location, the investigators found that that location had much less impact than did *level of survey experience*. SLCs in turn hired two interviewers each, by spreading word of the openings through their extensive networks in the Asian community. Korean interviewers were found to be especially difficult to recruit.

## 1.2. Instrument translation

Translation received considerable attention: As a quality control step, SLCs first reviewed the prior work of the initial questionnaire translators, and in order to reconcile differing opinions, Westat also hired a set of expert reviewers to arbitrate disagreements. The SLCs found a much greater need for revisions than had been anticipated; in reviewing the translated instruments, they each reported that the translations were literal, resulting in questions that were sometimes too wordy or even confusing (Chinese and Korean teams suggested revisions to about 60 percent of the approximately 200 survey items, and translation revisions were even more extensive for the Vietnamese version).

## 1.3. Materials preparation and interviewer training

The investigators spent a considerable amount of time educating the SLCs about survey development work, especially the purpose and uses of cognitive interviewing, in-depth discussions of how cultural issues may impact translation and data collection, recruiting techniques, supervising interviewers, and writing up results. The SLCs' main task related to the training was to translate the interview protocols into their individual languages. During this process, computer compatibility problems created a variety of unforeseen problems. Cutting and pasting from a document saved in one language into a document saved in another language resulted in a variety of computer problems, perhaps because questionnaire document travelled through at least four different computers, including NCI's, the translators', Westat's, and the SLCs' personal computers.

The training itself was conducted primarily in English, but interviewers and SLCs practiced the role plays in their individual languages. Overall, Westat staff felt this approach to be effective, but reported underestimating how difficult it is for those with no survey research background to grasp cognitive interview purposes, concepts, and techniques. *Most noticeably, interviewers and SLCs struggled to distinguish between survey items and cognitive interview probes.* In retrospect, it might have been better to increase training session to eight or ten hours, rather than six, and to equip interviewers with an in-depth understanding of why each survey item is being asked, as well as the purpose of each probe. Overall, however, the investigators felt that the best alternative would be to *have the SLCs themselves to conduct the interviews*, given that SLCs had much more exposure to survey methods and the survey itself over the course of several weeks as they prepared for testing. This observation is consistent with the feeling within U.S. cognitive laboratories that interviews are best conducted by a cadre of experienced, long-term employees, as opposed to lower-level interviewers who are simply trained as needed to conduct interviews.

## 1.4. Subject recruitment

SLCs initially expressed a variety of concerns related to recruiting and interviewing respondents: All thought a planned U.S. \$35 incentive would be too low to convince respondents to participate in an hour-long interview, and suggested the incentive be supplemented with small gifts such as fruit or cookies, particularly if the interview were held in the respondent's home. To find eligible respondents, the SLCs used a variety of methods, including flyers, newspaper advertisements, community events, word-of-mouth, and personal contacts. To one extent or another, these were effective, though more elaborate than the types required for usual (general, English-language) cognitive interviews. For example, some potential respondents suggested that newspaper ads are more trustworthy than flyers. In the end, personal contacts were by far the most effective tool for recruiting respondents, for all four SLCs. However, some of the SLCs were reluctant to use their personal networks at first, both because they didn't want to ask people about their smoking habits

(an apparently private subject in some Asian communities), and because they didn't want to "owe favors."

### **1.5. Conducting interviews**

All three SLCs oversaw completion of nine interviews, and five additional Vietnamese interviews were conducted in a second testing round. Given the language barrier, Westat team members were not able to monitor interviews. As an alternative quality control measure, SLCs participated in a telephone or in-person debriefing meeting after every second or third interview. After the first few interviews, SLCs reported that respondents were impatient with the redundancy of the interviewers' questions. The investigators determined that interviewers were administering the probes word-for-word, regardless of whether respondents had already provided the information the probes were designed to elicit. Further, interviewers were hesitant to deviate from probes, either by using their own words, or by following up unanticipated problems with unscripted probes.

The researchers initially had concerns about the consistency of cognitive interviewer behavior across languages, and to some extent problems in the area were observed. In particular, based on debriefing meetings with the Vietnamese SLC, the Westat team experienced doubts about the quality of Vietnamese interviews. At that time, the SLC suggested that Vietnamese speakers cannot understand true/false items because they are not posed as questions, and to remedy the problem, she recommended converting each true/false item into a question. However, a Vietnamese-speaking Westat reviewer who was independently consulted felt that the true/false statements were clear and understandable. In further reviewing the tapes, that reviewer discovered a variety of errors: the probes and interviewer instructions had been mis-translated, and the survey itself was administered improperly. For example, the series of true/false items were read with no pauses, so the respondent was not given an opportunity to answer each individual statement. Further, interviewer instructions and skip patterns were sometimes read aloud to the respondents; and survey items and probes were read using a monotone delivery, and probes were administered as if they were survey items rather than as needed. Because of these problems, the Vietnamese interviews were considered unusable, and five more were conducted. This finding speaks to the difficulty that English-only researchers may have in retaining appropriate control and monitoring of non-English cognitive interviews.

For future cognitive testing in foreign languages, Westat staff recommended dividing the first round of interviewing into two sub-rounds. The first sub-round would consist of three interviews, and after completion of these three, an expert reviewer would listen to the taped interviews and summarize any problems or issues interviewers encountered in administering either the survey instrument or the probes. If needed, a remedial training would take place before proceeding.

### **1.6. Problems identified through cognitive interviews**

To analyze interview results, SLCs wrote a detailed summary of each interview, using a note-taking template to guide them. For some items, detailed questions were asked to ensure that the issues of interest were addressed consistently across interviews and among the SLCs. Overall, the following categories of problems were found (with illustrative examples):

#### **A. Translation errors:**

(1) When hearing the question that asked whether they had smoked 100 cigarettes, most Chinese respondents answered with some version of, "In one day?" It was later determined that the phrase "in your entire life" had been left out of the translation.

(2) A Korean translation reversed the meaning of the response choices for the item: “Does anybody smoke cigarettes, cigars, or pipes inside your home?” Translated back to English, the Korean version read, in essence “Is there *not* anyone smoking cigarettes, cigars, or pipes anywhere inside of your home?” In Korean, the correct answer if no one smokes is “yes” (i.e., it is true that there is not anyone smoking cigarettes”).

(3) In Vietnamese, most respondents interpreted the word “community” (“In your opinion, how easy is it for minors to buy cigarettes and other tobacco products in your community?”) as the Vietnamese people in general, presumably because the translation of “community” carries a political connotation. The Vietnamese word for “neighborhood” was also problematic, as it implies that the respondent has a relationship with his/her neighbors. The authors recommended translating the question to emphasize physical location: “In your opinion, how easy is it for a minor to buy cigarettes or cigarette products in the area where you are now living?”

#### B. General problems (not culturally specific):

(1) When asked whether they would go to the store in a bad rainstorm for cigarettes, some Chinese subjects who said “no” indicated, upon probing, that they would never have to do this, because they always buy enough cigarettes to ensure they do not run out, or would simply borrow from friends until the weather cleared up. This did not appear to necessarily be associated with Chinese culture, however (and may relate more to the culture of smoking).

(2) Similar to previous (English-language) findings, some Chinese respondents did not think of themselves when answering whether anyone smokes inside their homes; the sponsor was encouraged to add “including yourself” or “including those who live here” to the question.

(3) As has generally been found to be the case, older respondents tended to have more difficulty answering questions, and required many to be repeated several times before they felt able to answer.

(4) Koreans, as members of other previously tested groups (Spanish, English), had little trouble understanding the question “What is the total number of years you have smoked every day?” until a follow-up instruction to exclude any time they stayed off cigarettes for six months or longer confused them.

(5) In keeping with another common finding, the Korean translation asks whether respondents have been to a health professional, but does not distinguish whether it was for their own health or someone else’s. Several respondents answered “yes” because they had accompanied others for a doctor visit (e.g., taking children to the pediatrician or one’s wife to the OB/GYN), even though they themselves had not seen a health professional in the past year. Project staff recommended adding to the question the phrase “for your own health...” (a similar recommendation has been made in the past for an English version of that question).

### C. Culturally-specific issues:

- (1) The difference between the words “adult” and “minor” is small enough in Chinese that the latter was often misheard as the former. The investigators suggested that interviewers be instructed to enunciate the word “minor” to avoid misunderstandings.
- (2) In Korean, The translation of “how soon” (“How soon after you wake up do you typically smoke your first cigarette of the day?”) involved a somewhat unusual Korean phrase. The suggested revision translates back to English as “After waking up on the mornings of days that you smoke, how long of a period of time goes by before you smoke?”
- (3) An item which asked whether respondents have ever switched from a stronger to a lighter cigarette posed difficulties for respondents who started smoking a Korean brand of cigarettes, then switched to an American brand, as cigarette manufacturers in Korea are not required to include information about tar and nicotine on each package.

## 1.7. Summary

The CPS-TUS study seems to have identified a range of problems, but involved a considerably greater amount of complexity, cost, and time than would cognitive interviewing of 32 English-speakers. To a great extent, the positive effects of testing appeared to be not so much culturally-specific cognitive testing results themselves (which were only moderate in scope), but rather improvements to the translation, and recognition of general (e.g., non-culturally specific) problems simply as a result of “more testing” being done.

The overall findings of the study were that:

- (a) Translation (of both questionnaires and cognitive protocols) is a major undertaking that requires significant resources;
- (b) Non-English interviewers need to be carefully developed, as opposed to hired and quickly trained; further, it is vital that cognitive interviewers have a grasp not only of language, but as well an appreciation for questionnaire design, survey intent, and flexible approaches to cognitive techniques.

## 2. Study 2: Systematic Analysis of Cognitive Interviewing Results Across English and Spanish

### 2.1. Overview

The project is described in more detail by Miller, Willis, Eason, Moses, and Canfield (under review). This study, overseen by staff of the National Center for Health Statistics, involved a very different approach to cross-cultural cognitive interviewing than Study 1 above. Rather than systematically documenting the *procedures* appropriate for the conduct of cross-cultural (and multi-lingual) cognitive interviewing, the investigators instead emphasized the manner in which cognitive testing *results are processed*, evaluated, and compared. They first noted that there is little agreement among practitioners regarding the standards or criteria appropriate for cognitive interviews (Snijders 2003), especially with respect to the nature of writing up results in a way that makes clear to investigators if problems exist. Therefore, they advocated supplementing the normally open-ended written cognitive interviews with quantifiable

outcome codes, especially for cross-cultural studies, where reliance on purely qualitative and sometime impressionistic interviewing approaches might lead to erroneous conclusions about cross-cultural discrepancies.

## **2.2. Procedure**

The investigators conducted sixty-seven cognitive interviews, divided equally between (a) (self-reported) Hispanic and (b) Non-Hispanic (both White and Black) participants, in both urban (Washington D.C) and rural/suburban locations (two areas of Northwest Ohio). The D.C. area interviews were conducted in the Questionnaire Design Research Laboratory at the National Center for Health Statistics. Ohio interviews were conducted either in the participant's home or in a private room of a community facility. Interviews of Hispanics were conducted by two bilingual consultants, one of whom had translation experience and had previously been trained in both questionnaire design and in cognitive interviewing techniques. Non-Hispanic (English language) interviews were conducted by several NCHS staff members, ranging in cognitive interviewing experience (from moderately to very experienced).

The interviews were based on an interviewer-administered health-survey questionnaire containing items covering chronic conditions, cancer screening, diet, physical activity and demographics. All but one of the cognitive interviews of Hispanics were conducted in Spanish, and all Non-Hispanics interviews were in English. The instrument and cognitive probe questions were translated from English to Spanish by one of the Spanish-speaking cognitive interviewers, with no further reconciliation or review. Spanish-language interviews were conducted by two interviewers who were trained for the activity; one of whom was an investigator, the other a college professor. Hence, both represented the approach advocated by Westat researchers, in which the cognitive interviewers are fairly high-level project investigators who have prior experience in questionnaire design and survey methods.

The cognitive interviews were semi-structured; along with the survey questions, the interview guide (protocol) consisted of several pre-scripted follow-up questions pertaining to participants' interpretations of key terms and overall comprehension of questions. These fixed probes ensured that this particular information was collected in every interview and could then be compared across all interviews. For a less standardized approach, interviewers were also instructed to exhibit latitude, and to inquire as to the ways in which participants constructed their answers to the survey questions, which further provides insight into potential sources of response error. These *emergent*, non-scripted probes were designed to help interviewers make sense of gaps or contradictions in participants' explanations, and to provide information needed to interpret question problems.

## **2.3. Problem codes**

"Problem codes," based on a standard question-response model (Comprehension, Retrieval, Decision and Response), were designed to indicate where subjects experienced definable problems with the tested questions (see Table 1).

**Table 1: Question-Response Problem Codes (Miller, et al.).**

Comprehension	1	Term: Subject does not understand or know the meaning of specific words
	2	Question: Subject does not understand the question as a whole because of vagueness or complexity
Retrieval	3	Subject does not know (and never knew) the requested information
	4	Subject is unable to remember requested information
Decision	5	Subject is unable to make calculations necessary to arrive at the answer
	6	Question sensitivity or perceived negative reaction by subject
	7	Subject is unable to decide on a response
	8	Subject is found to estimate either too high or too low
Response	9	Response categories do not match subject's internal representation of the answer
----	0	No problems observed

## 2.4. Results

Despite the development of multiple coding categories, the investigators determined that the most useful overall measure was one simply indicating if *any* type of problem was found in a particular interview, and that the (text-based) qualitative results were useful for further diagnosing the nature of the problem. To assess whether observed problems were systematically related to ethnicity or to other measured subject characteristics, cross-tabulations and logistic regression analyses were conducted for 18 questions, involving age, gender, income, education, and ethnicity. These analyses determined that for these items, ethnic group membership was the strongest overall predictor of problem code frequency, with Hispanics generally experiencing more difficulties than non-Hispanics (for 5 items), but with Hispanics seemingly having fewer problems for two other questions. Hispanics produced significantly more problems with a question on ever having cancer, and significantly less for one on combined household income. Five other questions exhibiting Hispanic/Non-Hispanic differences involved food and meal questions (see Table 2). Somewhat surprisingly, the analysis revealed that other demographic characteristics (gender, age, educational level, and income) had, at most, weak effects.

Qualitative analysis was used to facilitate interpretation of these results, and to pinpoint the character of the observed problem. From the start of Spanish language interviewing, it was clear that some translated survey questions caused interpretation difficulties for Hispanic subjects. That is, particular words were translated literally from English and, because of cultural differences, did not convey the same meaning. For example, the phrase *frijoles con chile* was intended to mean chili beans, but was interpreted by most Hispanic participants as beans with hot sauce. Additionally, some words varied by region (e.g., Puerto Rican Spanish uses *nami* for yam, while Mexican Spanish uses *camote*) or were inappropriately formal forms of Spanish (e.g., the word *fiambre* for lunchmeat). Consequently, these terms were not universally understood in the same manner by Spanish-speaking subjects. This type of translation problem seemed to account for the high percentage of Hispanics experiencing problems with the red meat question.

Similarly, some words in Spanish consisted of more than one meaning and could easily be misinterpreted, depending on the context in which they occurred. The word *comida* can mean either “meal,” “food,” or the name of a particular meal (e.g., the English word for dinner). Consequently, the question “*Did you eat a morning meal?*” was translated as “*¿Ayer comió Usted la comida de la mañana?*” but misunderstood by some Hispanic participants as “*Did you eat your dinner in the morning?*” This interpretive issue accounted for a significant degree of the ethnicity-based problems regarding the meal questions, and illustrates how cognitive probing brings out otherwise “silent misunderstandings,” as termed by DeMaio & Rothgeb (1996).

A qualitative examination of the interviews also revealed why Non-Hispanic subjects, in comparison to Hispanics, were likely to experience problems with a question on cooking oil. Many Hispanics reported using either butter or lard, but nothing else, and were able to provide a report concerning the one they used most often with little difficulty. Non-Hispanic participants were much more likely to cook with a variety of oils, and experienced trouble in reporting the most frequently used type.

Reiterating the results of Study 1, some findings were unrelated to language or culture. During the course of conducting the interviews, it became obvious that, for the question “*Yesterday did you eat any beans such as kidney beans, refried beans, chili beans, bean soup, bean salad or lentils?*,” participants adopted differing interpretations of the word *bean*. Some participants viewed the question as asking about legumes only, while others included any kind of bean, even green beans. Yet, at that point, interviewers could only speculate whether there was a particular group of participants using a specific interpretive pattern. It was hypothesized that older participants and perhaps less educated participants would be less inclined to view the question as asking about legume consumption. As it turned out, based on statistical (regression) analysis, the patterns of interpretation were not related to specific demographic group membership, as “green bean error” was found to be essentially random.

**Table 2: Miller et al. study: Percentage of participants having response problems, by ethnicity.**

Tested question:	Hispanics	Non-Hispanics
1) How many times did you eat red meat, including beef, pork, lamb, or lunchmeat, hot dogs or sausages made from beef, pork or lamb yesterday?	77.1% (27/35)	35.5% (11/31)
2) Did you eat a morning meal yesterday?	66.7% (24/36)	6.5% (2/31)
3) Did you eat a midday meal yesterday?	54.2% (13/24)	13.8% ( 4/29)
4) Did you eat an evening meal yesterday?	57.7% (15/26)	6.7% (2/30)
5) When you use butter or oils for cooking or preparing your food, which of the following types do you use most often? 1) Butter, Margarine, Lard or Shortening, 2) Olive oil or Canola oil, 3) Corn oil, Vegetable oil, Peanut oil, Soy oil, 4) non-stick spray, 5) Don't use fat	8.3% (3/36)	38.7% (12/31)

## 2.5. Conclusions, Implications, and Caveats

The Hispanic/Non-Hispanic study suggested that non-trivial differences may exist between these groups in responding to common health survey questions. As found by Kudela et al., some of these were due to translation, some to cultural influences. Further, some outcomes “favored” Non-Hispanics, others Hispanics. Concerning the basic research question – whether systematic coding of results enhances cross-cultural interpretation – this looked promising, but the investigators recognized several limitations and caveats. From an operational point of view, systematic coding of cognitive interviewing results may not always be feasible. In many pretesting studies (e.g., that in Study 1), much smaller rounds of interviews are generally conducted than were included in the current study; restrictions by the U.S. Office of Management and Budget often limit the size of an interviewing round to no more than nine subjects. In such cases strict quantification of results will not supply information sufficient to identify cross-cultural differences. Further, the authors admit that the amount of time required to ensure that all codes were applied consistently across all interviews, and to develop an analyzable data set, was considerable. *Therefore, as Kudela et al. they cite the need for significant resources in cross-cultural cognitive interviewing studies, but for different reasons.*



Finally, and most importantly, it remains the case that even though cognitive interviewing results were systematically coded and therefore produce quantitative, statistically-analyzable data, these data are only as good as the information they derive from. To the extent that the Spanish interviewers were behaving differently from the English-speaking ones (unknown in this case), it is unclear whether significant differences in coding frequency between cultural groups reveal variation that is due to ethnicity/language, or whether this simply represents interviewer variation. On the other hand, the fact that the qualitative results appear coherent, in the context of the qualitative results, does tend to obviate (though not to eliminate) this concern.

### **3. Overall recommendations based on Study 1 and Study 2**

Although the Westat and NCHS studies were somewhat dissimilar, involving different populations, approaches, and even research questions, they point to similar conclusions and implications, concerning the conduct of cross-cultural cognitive interviewing:

#### **3.1. Cross-cultural studies require an exponential increase in resources**

This may not be a universal truth that will apply for all time, as means for the conduct of cognitive interviewing like will become more efficient and cost-effective. Initially, however, investigators need to realize the additional effort is necessary in the areas of instrument and protocol translation and verification, subject location and recruitment, interviewer training, coordination, analysis, and interpretation of results, and overall coordination of a complex project.

#### **3.2. Some of the positive effects observed are simply a function of additional testing**

In both studies, some problems seemed culturally non-specific (that is, “etic” rather than “emic,” in anthropological terms); they appear to simply be problems with survey questions of the type normally identified through cognitive interviewing techniques. In cases where such problems had been previously identified, this is reassuring, in that it serves to verify that cognitive techniques have been applied consistently across group, and produce reliable results. With respect to novel, unanticipated problems that still appear to apply across evaluated groups, it may be the case that cross-cultural testing only serves as additional (but useful) testing, where issues of language and culture are somewhat irrelevant.

#### **3.3. Issues of staff selection and training are paramount**

It is increasingly apparent that cognitive interviewing is an acquired skill, and that some procedures and interviewer behaviors that seem obvious to seasoned practitioners (e.g., avoid biased probing, be selective in the administration of probes, recognize when a subjects’ comment indicates a problem with the tested question) are not at all clear to beginners. Hence, significant resources need to again be devoted to developing a cadre of qualified, bilingual cognitive interviewers. Optimally, once these individuals are trained, they will be available to assist with later investigations.

#### **3.4. Cognitive protocols for cross-cultural studies may benefit from increased standardization**

Especially if novice interviewers are used for cross-cultural studies (and perhaps generally), we advise that the cognitive protocol be very structured in nature, in order to minimize idiosyncratic approaches or departures from acceptable practice. The “frontloading” of probes in this manner does come at a cost, as this sacrifices the opportunity to rely on “reactive” or “emergent” probes

that are used to investigate unanticipated problems (Willis, 2005). On balance, however, it may be worthwhile to constrain behaviors that could produce observed between-group differences that are simply artifacts of interviewing style.

### **3.5. The overseeing staff must create and maintain ways to stay as integrated into the process as possible**

Monolingual cognitive interviewing staff who oversee cross-cultural projects sometimes feel partly blind, as they cannot observe or directly evaluate cognitive interviews conducted in another language. Therefore, it is advisable to develop substitute mechanisms to ensure adequate monitoring and quality control. Interesting, Westat staff have recently reported having success with the practice of simultaneous translation of the cognitive interview, in which a Spanish-language interaction between the interviewer and respondent was directly observed (within the same room) by a language translator and the (English-speaking) project director. As the interview unfolded and was translated to the director, she was able to intervene when necessary to direct follow-up probing, or to otherwise re-direct the activities of the cognitive interviewer. Although this somewhat complex arrangement appeared to be workable, it may not function for all cultures.

However, it does well illustrate the value of openness to new approaches within the cross-cultural domain, and is one of many issues to be further investigated in this realm.

## **References**

- Behling, O., & Law, K.S. (2000). Translating questionnaires and other research instruments: Problems and solutions. London: Sage.
- Cannell, C. F., Fowler, F. J., & Marquis, K. (1968). The influence of interviewer and respondent psychological and behavioral variables on the reporting in household interviews. *Vital and Health Statistics*, Series 2, No. 26. Washington, DC: U.S. Government Printing Office.
- de la Puente, M., & Pan, Y. (2003). An overview of proposed Census Bureau Guidelines for the translation of data collection instruments and supporting materials. *Paper presented at the meeting of the Federal Committee on Statistical Methodology*, Arlington, VA.
- DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 177–195). San Francisco: Jossey-Bass.
- Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify problems with survey questions. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. (pp. 15-36) San Francisco: Jossey-Bass.
- Gerber, (1999). The view from anthropology: Ethnography and the cognitive interview. In M. Sirken, D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau (Eds.), *Cognition and Survey Research*.(pp 217-234). New York: Wiley.
- Harkness, J.A., Van de Vijver, F.J.R., & Mohler, P. (2003). Cross-Cultural Survey Methods. Hoboken, N.J.: Wiley.
- Ji, L.J., Schwarz, N., & Nisbett, R.E. (2000). Culture, autobiographical memory, and behavioral frequency reports: Measurement issues in cross-cultural studies. *Personality and Social Psychology Bulletin* 26, 586-594.

- Johnson, T.P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA – Nachrichten Spezial*, 3, 1-39.
- Kudela, M.S., Kevin, K., Tseng, M, Hum, M., Lee, S., Wong, C., McNutt, S., & Lawrence, D. (2003). Tobacco Use Cessation Supplement to the Current Population Survey Chinese, Korean, and Vietnamese Translations: Results of Cognitive Testing. Final Report submitted to the National Cancer Institute, Rockville, MD.
- McKay, R.B., Breslow, M.J., Sangster, R.L., Gabbard, S.M., Reynolds, R.W., Nakamoto, J.M., & Tarnai, J. (1996). Translating survey questionnaires: Lessons learned. *New Directions for Evaluation*, 70, 93-105.
- Miller, Willis, Eason, Moses, and Canfield (under review). Interpreting the Results of Cross-Cultural Cognitive Interviews: A Mixed-Method Approach.
- Pan, Y. (2003). The role of sociolinguistics in the development and conduct of Federal surveys. *Paper presented at the meeting of the Federal Committee on Statistical Methodology*, Arlington, VA.
- Rose, D., & de la Puente, M. (2003). Guideline implementation: Translating a new Canadian/U.S. health survey into Spanish. *Paper presented at the meeting of the Federal Committee on Statistical Methodology*, Arlington, VA.
- Snijders, G. (2003). Cognitive laboratory experiences: On pretesting computerized questionnaires and data quality. Unpublished Doctoral Dissertation, Utrecht University, Netherlands.
- Stewart, A.L. & Napoles-Springer, A. (2000). Health-related quality-of-life assessments in diverse population groups in the United States. *Medical Care*, 38, II-102—II-124.
- Warnecke, R.B., Johnson, T.P., Chavez, N., Sudman, S, O'Rourke, D.P., Lacey, L, & Horm, J. (1997). Improving question wording in surveys of culturally diverse populations. *Annals of Epidemiology*, 7, 334-342.
- Willis, G.B. (2005). Cognitive Interviewing: A Tool for Improving Questionnaire Design. Sage: Thousand Oaks, CA.